Optimising radio interferometric imaging with compressive sensing

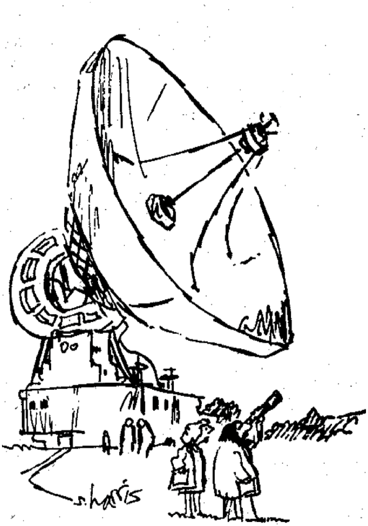
Jason McEwen
www.jasonmcewen.org
@jasonmcewen

Mullard Space Science Laboratory (MSSL)
University College London (UCL)

In collaboration with Laura Wolz, Filipe Abdalla, Rafael Carrillo & Yves Wiaux

Experimental Design and Big Data, University of Warwick, May 2015

Radio telescopes are big!



"Just checking."

Radio telescopes are big!

Radio interferometric telescopes

Next-generation of radio interferometry rapidly approaching

- Square Kilometre Array (SKA) construction scheduled to begin 2018.
- Many other pathfinder telescopes under construction, *e.g.* LOFAR, ASKAP, MeerKAT, MWA.
- Broad range of science goals.
- New modelling and imaging techniques required to ensure the next-generation of interferometric telescopes reach their full potential.

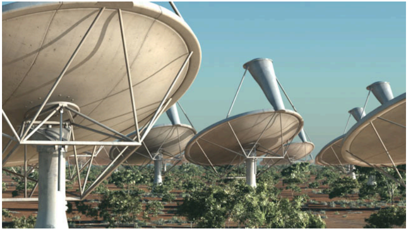


Figure: Artist impression of SKA dishes. [Credit: SKA Organisation]

Next-generation of radio interferometry rapidly approaching

- Square Kilometre Array (SKA) construction scheduled to begin 2018.
- Many other pathfinder telescopes under construction, *e.g.* LOFAR, ASKAP, MeerKAT, MWA.
- Broad range of science goals.
- New modelling and imaging techniques required to ensure the next-generation of interferometric telescopes reach their full potential.

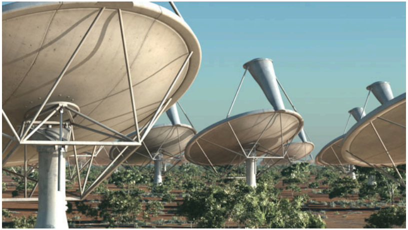
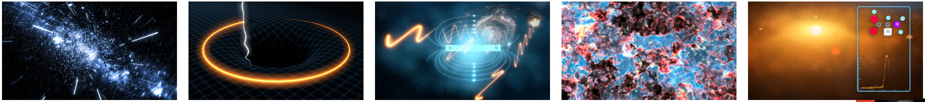


Figure: Artist impression of SKA dishes. [Credit: SKA Organisation]



(a) Dark-energy (b) GR (c) Cosmic magnetism (d) Epoch of reionization (e) Exoplanets

Figure: SKA science goals. [Credit: SKA Organisation]

Next-generation of radio interferometry rapidly approaching

- Square Kilometre Array (SKA) construction scheduled to begin 2018.
- Many other pathfinder telescopes under construction, *e.g.* LOFAR, ASKAP, MeerKAT, MWA.
- Broad range of science goals.
- New modelling and imaging techniques required to ensure the next-generation of interferometric telescopes reach their full potential.

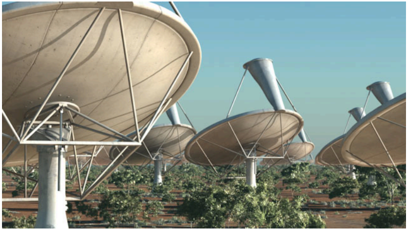
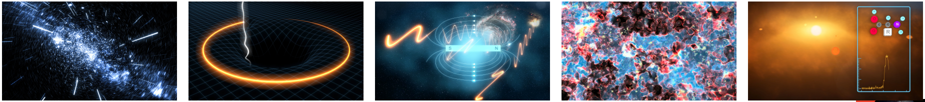


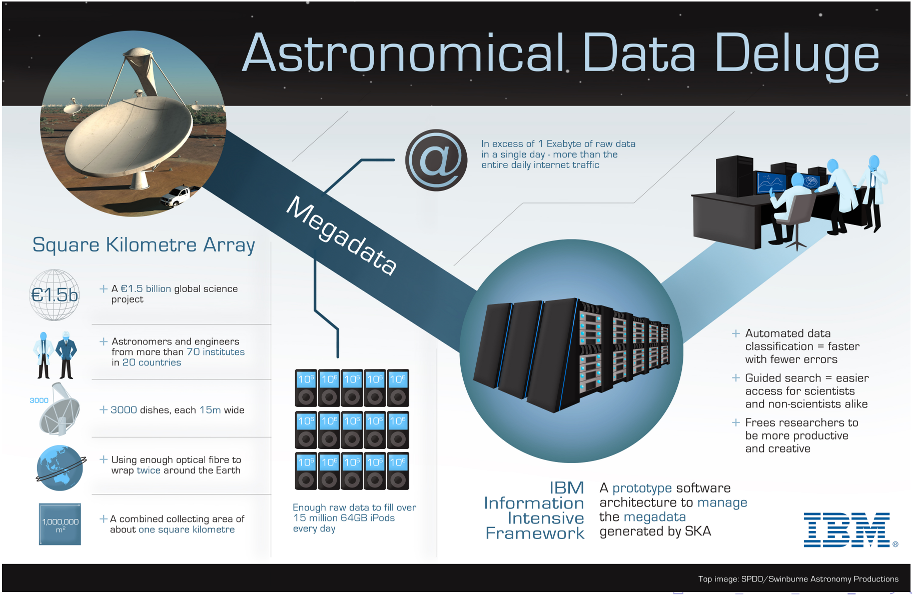
Figure: Artist impression of SKA dishes. [Credit: SKA Organisation]



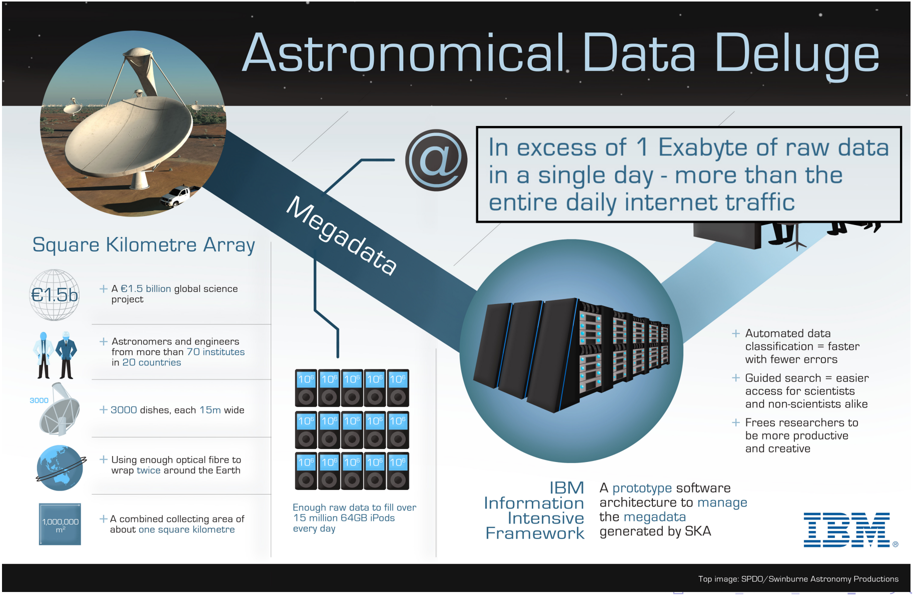
(a) Dark-energy (b) GR (c) Cosmic magnetism (d) Epoch of reionization (e) Exoplanets

Figure: SKA science goals. [Credit: SKA Organisation]

The SKA poses a considerable big-data challenge



The SKA poses a considerable big-data challenge



Outline

1. Compressive sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations
2. Interferometric imaging with compressive sensing
   - Imaging
   - SARA
   - Continuous visibilities
3. Optimising telescope configurations
   - Spread spectrum effect
   - Simulations

Outline

1. Compressive sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations

2. Interferometric imaging with compressive sensing
   - Imaging
   - SARA
   - Continuous visibilities

3. Optimising telescope configurations
   - Spread spectrum effect
   - Simulations

Compressive sensing

"Nothing short of revolutionary."

– National Science Foundation

- Developed by Candes *et al.* 2006 and Donoho 2006 (and others).
- Although many underlying ideas around for a long time.
- Exploits the sparsity of natural signals.

(a) Emmanuel Candes

(b) David Donoho

Compressive sensing

- Mystery of JPEG compression.
- Move compression to the acquisition stage → compressive sensing.
- Acquisition versus imaging.

Figure: TIFF (uncompressed) vs JPEG (compressed)

Compressive sensing

- Mystery of JPEG compression.
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.

Figure: TIFF (uncompressed) vs JPEG (compressed)

Compressive sensing

- Mystery of JPEG compression.
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.

Figure: TIFF (uncompressed) vs JPEG (compressed)

Compressive sensing

- Mystery of JPEG compression.
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.

Figure: TIFF (uncompressed) vs JPEG (compressed)

An introduction to compressive sensing

Operator description

- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - \Phi_0 - \\ - \Phi_1 - \\ \vdots \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$

An introduction to compressive sensing

Operator description

- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - & \Phi_0 & - \\ - & \Phi_1 & - \\ & \vdots & \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$

An introduction to compressive sensing

Operator description

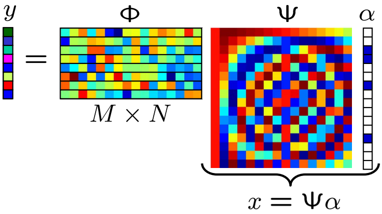
- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - \ \Phi_0 \ - \\ - \ \Phi_1 \ - \\ \vdots \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$



An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n}}\,.$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_0 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon}\ ,$$

  where the signal is synthesising by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solving this problem is difficult (combinatorial).

- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon}\ .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n}} .$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_0 \ \text{such that} \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon} ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon} .$$

An introduction to compressive sensing

Union of subspaces

- Space of sparse vectors given by the union of subspaces aligned with the coordinate axes.

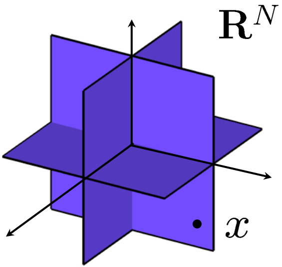


Figure: Space of the sparse vectors [Credit: Baraniuk]

An introduction to compressive sensing

Restricted isometry property (RIP)

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Restricted isometry property (RIP):

$$(1-\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \leq \|\Theta\boldsymbol{x}_1-\Theta\boldsymbol{x}_2\|_2^2 \leq (1+\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \,,$$

  for K-sparse $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, where $\Theta = \Phi\Psi$.
- Measurement must preserve geometry of sets of sparse vectors.

An introduction to compressive sensing

Restricted isometry property (RIP)

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Restricted isometry property (RIP):

$$(1-\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \leq \|\Theta\boldsymbol{x}_1-\Theta\boldsymbol{x}_2\|_2^2 \leq (1+\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \,,$$

  for K-sparse $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, where $\Theta = \Phi\Psi$.
- Measurement must preserve geometry of sets of sparse vectors.

An introduction to compressive sensing

Restricted isometry property (RIP)

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Restricted isometry property (RIP):

$$(1-\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \leq \|\Theta\boldsymbol{x}_1-\Theta\boldsymbol{x}_2\|_2^2 \leq (1+\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \,,$$

  for K-sparse $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, where $\Theta = \Phi\Psi$.
- Measurement must preserve geometry of sets of sparse vectors.

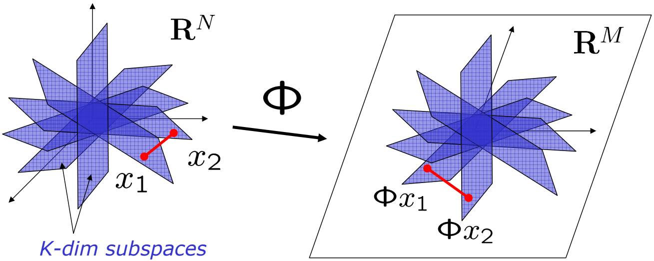


Figure: Measurement must preserve geometry of sets of sparse vectors. [Credit: Baraniuk]

An introduction to compressive sensing

Intuition

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Geometry of ℓ_0, ℓ_2 and ℓ_1 problems.

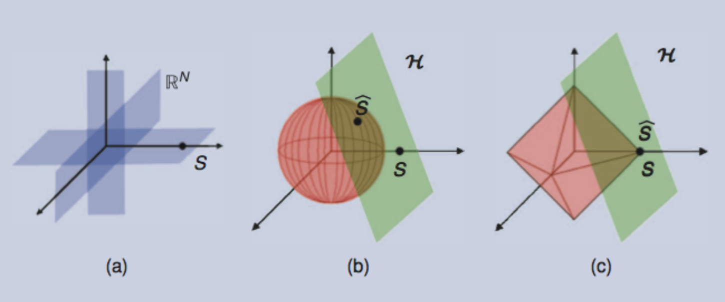


Figure: Geometry of (a) ℓ_0 (b) ℓ_2 and (c) ℓ_1 problems. [Credit: Baraniuk (2007)]

An introduction to compressive sensing

Coherence

- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$

- Robust to noise.

An introduction to compressive sensing

Coherence

- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$

- Robust to noise.

An introduction to compressive sensing
Coherence

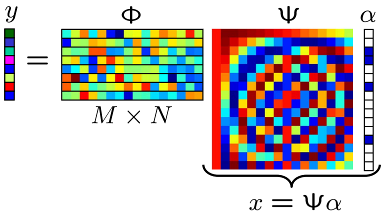
- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$



- Robust to noise.

An introduction to compressive sensing

Coherence

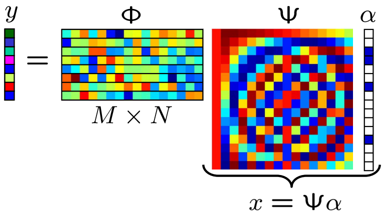
- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$



- Robust to noise.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$x^\star = \arg\min_{x} \|\Omega x\|_1 \text{ such that } \|y - \Phi x\|_2 \leq \epsilon .$$

analysis

- Contrast with synthesis-based approach:

$$x^\star = \Psi \cdot \arg\min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\arg\min} \, \|\Omega \boldsymbol{x}\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min} \, \|\boldsymbol{\alpha}\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \,.$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \,.$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \,.$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \,.$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

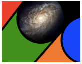
analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.



Analysis vs synthesis
Comparison

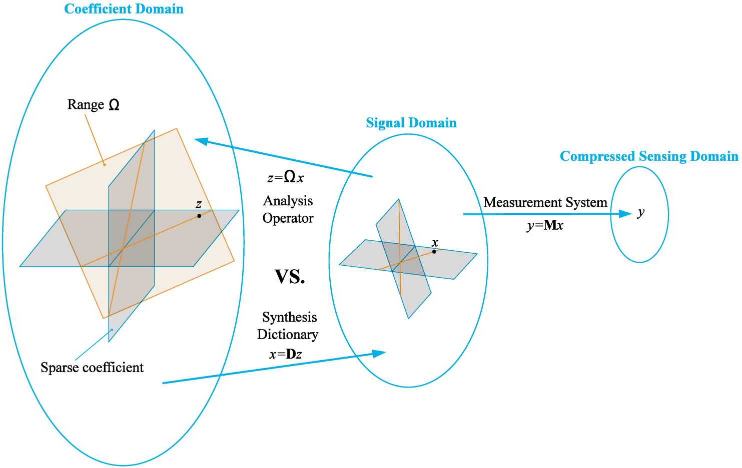


Figure: Analysis- and synthesis-based approaches [Credit: Nam *et al.* (2012)].

Analysis vs synthesis

Comparison

- Synthesis-based approach is more general, while analysis-based approach more restrictive.
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).

Analysis vs synthesis

Comparison

- Synthesis-based approach is more general, while analysis-based approach more restrictive.
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).

Analysis vs synthesis

Comparison

- Synthesis-based approach is more general, while analysis-based approach more restrictive.
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \Phi\Psi\boldsymbol{\alpha} + \boldsymbol{n} \ .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y}\,|\,\boldsymbol{\alpha}) \propto \exp\Big(\|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2/(2\sigma^2)\Big) \ .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Big(-\beta\|\boldsymbol{\alpha}\|_1\Big) \ .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^\star_{\text{MAP-Synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\ \mathrm{P}(\boldsymbol{\alpha}\,|\,\boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \ .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$y = \Phi\Psi\boldsymbol{\alpha} + \boldsymbol{n} \ .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{\alpha}) \propto \exp\Big(\|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 / (2\sigma^2) \Big) \ .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Big(-\beta \|\boldsymbol{\alpha}\|_1 \Big) \ .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^{\star}_{\text{MAP-Synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\, \mathrm{P}(\boldsymbol{\alpha} \mid \boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \ .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} \, .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{\alpha}) \propto \exp\Big(\| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 / (2\sigma^2) \Big) \, .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Big(-\beta \| \boldsymbol{\alpha} \|_1 \Big) \, .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^\star_{\text{MAP-Synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\, \mathrm{P}(\boldsymbol{\alpha} \,|\, \boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 + \lambda \| \boldsymbol{\alpha} \|_1 \, .$$

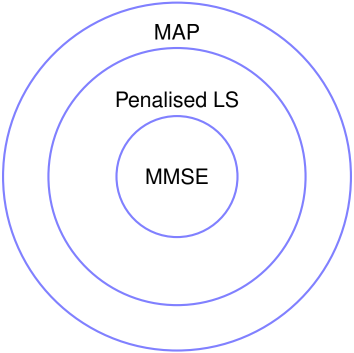
synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

Other Bayesian interpretations of the synthesis-based approach

- Other Bayesian interpretations are also possible (Gribonval 2011).
- Minimum mean square error (MMSE) estimators
  - $\subset$ synthesis-based estimators with appropriate penalty function, *i.e.* penalised least-squares (LS)
  - $\subset$ MAP estimators



Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\Omega\boldsymbol{x}\|_1 \ .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^{\dagger}$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \Omega^{\dagger} \cdot \arg\min_{\boldsymbol{\gamma} \in \text{column space } \Omega} \|\boldsymbol{y} - \Phi\Omega^{\dagger}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \ .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\Omega\boldsymbol{x}\|_1 \ .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\boldsymbol{\gamma} \in \text{column space } \Omega} \|\boldsymbol{y} - \Phi\Omega^\dagger\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \ .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$x^\star_{\text{MAP-Analysis}} = \arg\max_{x} \mathrm{P}(x \mid y) = \arg\min_{x} \|y - \Phi x\|_2^2 + \lambda \|\Omega x\|_1 \ .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$x^\star_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\gamma \in \text{column space } \Omega} \|y - \Phi\Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 \ .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2^2 + \lambda \|\Omega \boldsymbol{x}\|_1 \ .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\boldsymbol{\gamma} \in \text{column space } \Omega} \|\boldsymbol{y} - \Phi \Omega^\dagger \boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \ .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Outline

Radio interferometric inverse problem

- The complex visibility measured by an interferometer is given by

$$y(\boldsymbol{u}, w) = \int_{D^2} A(\boldsymbol{l})\, x(\boldsymbol{l})\, C(\|\boldsymbol{l}\|_2)\, \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{u}\cdot\boldsymbol{l}}\, \frac{\mathrm{d}^2\boldsymbol{l}}{n(\boldsymbol{l})},$$

visibilities

where the w-modulation $C(\|\boldsymbol{l}\|_2)$ is given by

$$C(\|\boldsymbol{l}\|_2) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)}.$$

w-modulation

- Various assumptions are often made regarding the size of the field-of-view:
  - Small-field with $\|\boldsymbol{l}\|^2 w \ll 1$ $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) \simeq 1$
  - Small-field with $\|\boldsymbol{l}\|^4 w \ll 1$ $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) \simeq \mathrm{e}^{\mathrm{i}\pi w\|\boldsymbol{l}\|^2}$
  - Wide-field $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) = \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)}$

Radio interferometric inverse problem

- The complex visibility measured by an interferometer is given by

$$y(\boldsymbol{u}, w) = \int_{D^2} A(\boldsymbol{l})\, x(\boldsymbol{l})\, C(\|\boldsymbol{l}\|_2)\, \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{u}\cdot\boldsymbol{l}} \frac{\mathrm{d}^2\boldsymbol{l}}{n(\boldsymbol{l})},$$

visibilities

where the w-modulation $C(\|\boldsymbol{l}\|_2)$ is given by

$$C(\|\boldsymbol{l}\|_2) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)}.$$

w-modulation

- Various assumptions are often made regarding the size of the field-of-view:
  - Small-field with $\|\boldsymbol{l}\|^2\, w \ll 1$ $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) \simeq 1$
  - Small-field with $\|\boldsymbol{l}\|^4\, w \ll 1$ $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) \simeq \mathrm{e}^{\mathrm{i}\pi w \|\boldsymbol{l}\|^2}$
  - Wide-field $\Rightarrow$ $C(\|\boldsymbol{l}\|_2) = \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)}$

Radio interferometric inverse problem

- The complex visibility measured by an interferometer is given by

$$y(\boldsymbol{u}, w) = \int_{D^2} A(\boldsymbol{l})\, x(\boldsymbol{l})\, C(\|\boldsymbol{l}\|_2)\, \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{u}\cdot\boldsymbol{l}}\, \frac{\mathrm{d}^2\boldsymbol{l}}{n(\boldsymbol{l})} ,$$

visibilities

where the w-modulation $C(\|\boldsymbol{l}\|_2)$ is given by

$$C(\|\boldsymbol{l}\|_2) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)} .$$

w-modulation

- Various assumptions are often made regarding the size of the field-of-view:
  - Small-field with $\|\boldsymbol{l}\|^2\, w \ll 1 \quad \Rightarrow \quad C(\|\boldsymbol{l}\|_2) \simeq 1$
  - Small-field with $\|\boldsymbol{l}\|^4\, w \ll 1 \quad \Rightarrow \quad C(\|\boldsymbol{l}\|_2) \simeq \mathrm{e}^{\mathrm{i}\pi w \|\boldsymbol{l}\|^2}$
  - Wide-field $\quad \Rightarrow \quad C(\|\boldsymbol{l}\|_2) = \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-\|\boldsymbol{l}\|^2}\right)}$

Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} ,$$

  where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} ,$$

where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} ,$$

  where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.

Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \, ,$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis Pursuit (BP) denoising problem

$$\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\boldsymbol{\alpha}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon \, , \quad \text{BPDN}$$

where the image is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Total Variation (TV) denoising problem

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\arg\min} \|\boldsymbol{x}\|_{\text{TV}} \ \text{such that} \ \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2 \leq \epsilon \, . \quad \text{TVDN}$$

Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} ,$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis Pursuit (BP) denoising problem

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon , \quad \text{BPDN}$$

where the image is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Total Variation (TV) denoising problem

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\text{TV}} \ \text{such that} \ \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2 \leq \epsilon . \quad \text{TVDN}$$

Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \; ,$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis Pursuit (BP) denoising problem

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon \, ,$$

BPDN

where the image is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Total Variation (TV) denoising problem

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathrm{TV}} \ \text{such that} \ \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2 \leq \epsilon \, .$$

TVDN

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight.
  $\Rightarrow$ concatenation of 9 bases
- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T\bar{x}\|_1 \quad \text{subject to} \quad \|y - \Phi\bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0\,,$$

SARA

where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight.
$\Rightarrow$ concatenation of 9 bases
- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T\bar{x}\|_1 \quad \text{subject to} \quad \|y - \Phi\bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0\,,$$

SARA

where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2012)

- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

  thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight.
  $\Rightarrow$ concatenation of 9 bases

- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W\Psi^T\bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi\bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0\,,$$

  SARA

  where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging
Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2012)

- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \dots, \Psi_q],$$

  thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight.
  $\Rightarrow$ concatenation of 9 bases

- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

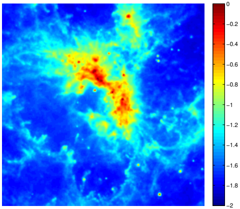
$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W \Psi^T \bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi \bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0 \,, \qquad \text{SARA}$$

  where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

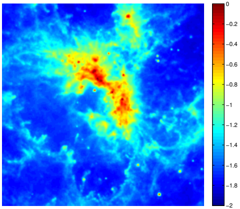
Results on simulations



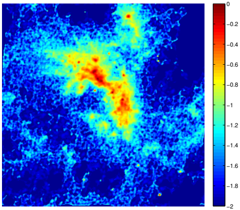
(a) Original

SARA for radio interferometric imaging

Results on simulations



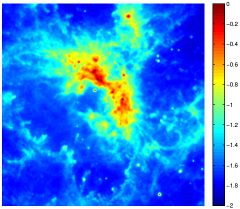
(a) Original



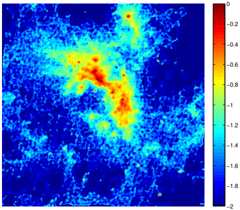
(b) “CLEAN” (SNR=16.67 dB)

SARA for radio interferometric imaging

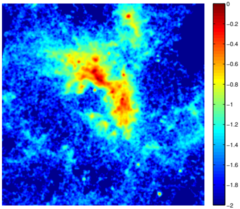
Results on simulations



(a) Original



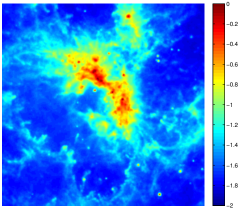
(b) “CLEAN” (SNR=16.67 dB)



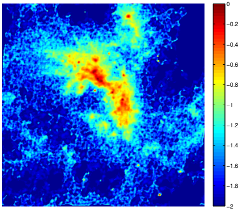
(c) “MS-CLEAN” (SNR=17.87 dB)

SARA for radio interferometric imaging

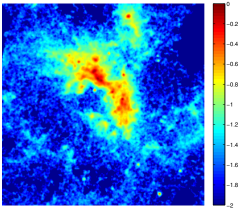
Results on simulations



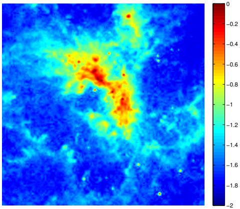
(a) Original



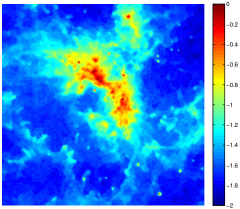
(b) “CLEAN” (SNR=16.67 dB)



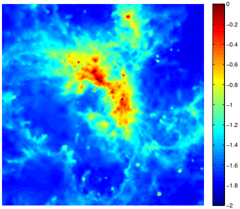
(c) “MS-CLEAN” (SNR=17.87 dB)



(d) BPDb8 (SNR=24.53 dB)



(e) TV (SNR=26.47 dB)



(f) SARA (SNR=29.08 dB)

SARA for radio interferometric imaging

Results on simulations

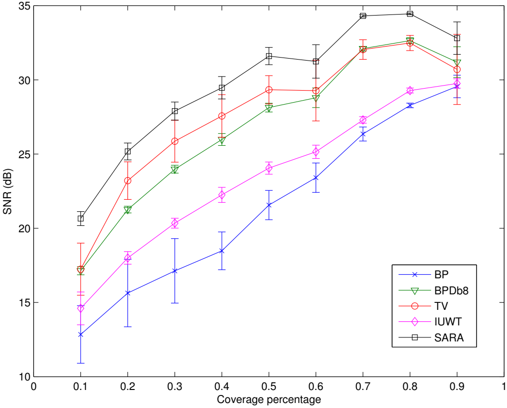


Figure: Reconstruction fidelity vs visibility coverage.

Supporting continuous visibilities

Algorithm

- Ideally we would like to model the continuous Fourier transform operator

$$\Phi = \mathbf{F}^{\mathrm{c}} \, .$$

- But this is impractically slow!
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

$$\Phi = \mathbf{G}\,\mathbf{F}\,\mathbf{D}\,\mathbf{Z} \, ,$$

where we incorporate:
  - convolutional gridding operator $\mathbf{G}$;
  - fast Fourier transform $\mathbf{F}$;
  - normalisation operator $\mathbf{D}$ to undo the convolution gridding;
  - zero-padding operator $\mathbf{Z}$ to upsample the discrete visibility space.

Supporting continuous visibilities

Algorithm

- Ideally we would like to model the continuous Fourier transform operator

$$\Phi = \mathbf{F}^{c} \, .$$

- But this is impractically slow!
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

$$\Phi = \mathbf{G}\,\mathbf{F}\,\mathbf{D}\,\mathbf{Z} \, ,$$

where we incorporate:

- convolutional gridding operator $\mathbf{G}$;
- fast Fourier transform $\mathbf{F}$;
- normalisation operator $\mathbf{D}$ to undo the convolution gridding;
- zero-padding operator $\mathbf{Z}$ to upsample the discrete visibility space.

Supporting continuous visibilities

Algorithm

- Ideally we would like to model the continuous Fourier transform operator

$$\boxed{\Phi = \mathbf{F}^{\mathrm{c}}} .$$

- But this is impractically slow!
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

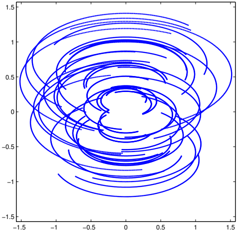
$$\boxed{\Phi = \mathbf{G}\,\mathbf{F}\,\mathbf{D}\,\mathbf{Z}} ,$$

where we incorporate:

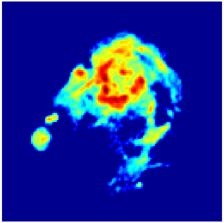
- convolutional gridding operator $\mathbf{G}$;
- fast Fourier transform $\mathbf{F}$;
- normalisation operator $\mathbf{D}$ to undo the convolution gridding;
- zero-padding operator $\mathbf{Z}$ to upsample the discrete visibility space.

Supporting continuous visibilities

Results on simulations



(a) Coverage

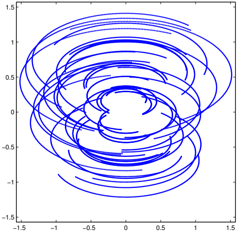


(b) M31 (ground truth)

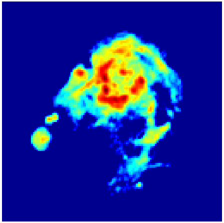
Figure: Reconstructed images from continuous visibilities.

Supporting continuous visibilities

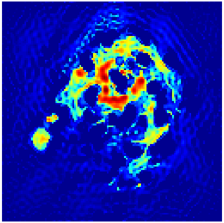
Results on simulations



(a) Coverage



(b) M31 (ground truth)

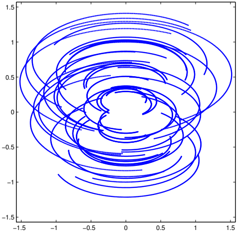


(c) “CLEAN” (SNR= 8.2dB)

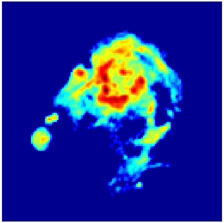
Figure: Reconstructed images from continuous visibilities.

Supporting continuous visibilities

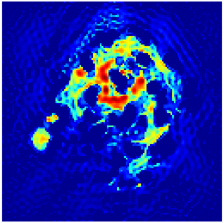
Results on simulations



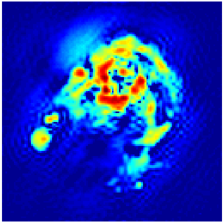
(a) Coverage



(b) M31 (ground truth)



(c) "CLEAN" (SNR= 8.2dB)

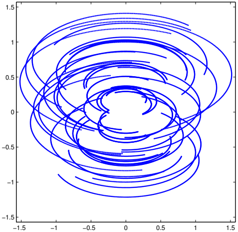


(d) "MS-CLEAN" (SNR= 11.1dB)

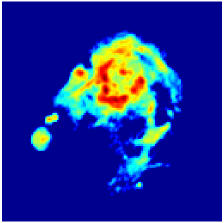
Figure: Reconstructed images from continuous visibilities.

Supporting continuous visibilities

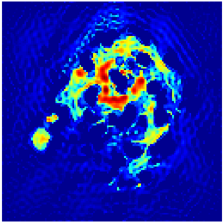
Results on simulations



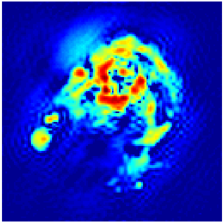
(a) Coverage



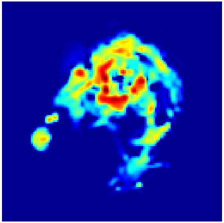
(b) M31 (ground truth)



(c) “CLEAN” (SNR= 8.2dB)



(d) “MS-CLEAN” (SNR= 11.1dB)



(e) SARA (SNR= 13.4dB)

Figure: Reconstructed images from continuous visibilities.

Outline

1. Compressive sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations
2. Interferometric imaging with compressive sensing
   - Imaging
   - SARA
   - Continuous visibilities
3. Optimising telescope configurations
   - Spread spectrum effect
   - Simulations

Optimising telescope configurations

Spread spectrum effect

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv e^{i2\pi w\left(1-\sqrt{1-l^2-m^2}\right)} \simeq e^{i\pi w \|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4 w \ll 1 ,$$

giving rise to to a linear chirp.

Optimising telescope configurations
Spread spectrum effect

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-l^2-m^2}\right)} \simeq \mathrm{e}^{\mathrm{i}\pi w \|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4 w \ll 1 ,$$

giving rise to to a linear chirp.

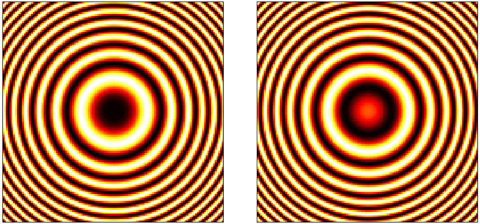
Optimising telescope configurations

Spread spectrum effect

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv e^{i 2\pi w \left(1-\sqrt{1-l^2-m^2}\right)} \simeq e^{i\pi w \|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4\, w \ll 1 ,$$

giving rise to to a linear chirp.



(a) Real part (b) Imaginary part

Figure: Chirp modulation.

Recap compressive sensing preliminaries

Sparsity and coherence

- What drives the quality of compressive sensing reconstruction?
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N \,,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| \,.$$

Recap compressive sensing preliminaries

Sparsity and coherence

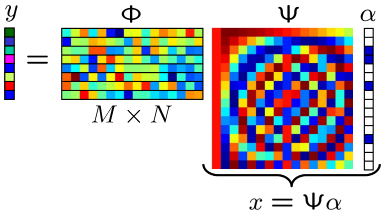
- What drives the quality of compressive sensing reconstruction?
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$



Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect
Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Overview

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013).

Spread spectrum effect

Sparse w-projection algorithm

- Apply the w-projection algorithm (Cornwell *et al.* 2008) to shift the w-modulation through the Fourier transform:

$$\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\,\mathbf{F}\,\mathbf{A} \; .$$

- Naively, expressing the application of the w-modulation in this manner is computationally less efficient that the original formulation but it has two important advantages.
- Different w for each (u, v), while still exploiting FFT.
- Many of the elements of $\hat{\mathbf{C}}$ will be close to zero.
- Support of w-modulation in Fourier space determined dynamically.

Spread spectrum effect

Sparse w-projection algorithm

- Apply the w-projection algorithm (Cornwell *et al.* 2008) to shift the w-modulation through the Fourier transform:

$$\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\,\mathbf{F}\,\mathbf{A} .$$

- Naively, expressing the application of the w-modulation in this manner is computationally less efficient that the original formulation but it has two important advantages.
- Different w for each (u, v), while still exploiting FFT.
- Many of the elements of $\hat{\mathbf{C}}$ will be close to zero.
- Support of w-modulation in Fourier space determined dynamically.

Spread spectrum effect

Sparse w-projection algorithm

- Apply the w-projection algorithm (Cornwell *et al.* 2008) to shift the w-modulation through the Fourier transform:

$$\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\,\mathbf{F}\,\mathbf{A} \; .$$

- Naively, expressing the application of the w-modulation in this manner is computationally less efficient that the original formulation but it has two important advantages.
- Different w for each (u, v), while still exploiting FFT.
- Many of the elements of $\hat{\mathbf{C}}$ will be close to zero.
- Support of w-modulation in Fourier space determined dynamically.

Spread spectrum effect

Sparse w-projection algorithm

- Apply the w-projection algorithm (Cornwell *et al.* 2008) to shift the w-modulation through the Fourier transform:

$$\boxed{\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\,\mathbf{F}\,\mathbf{A}}\ .$$

- Naively, expressing the application of the w-modulation in this manner is computationally less efficient that the original formulation but it has two important advantages.
- Different w for each (u, v), while still exploiting FFT.
- Many of the elements of $\hat{\mathbf{C}}$ will be close to zero.
- Support of w-modulation in Fourier space determined dynamically.

Spread spectrum effect

Sparse w-projection algorithm

- Apply the w-projection algorithm (Cornwell *et al.* 2008) to shift the w-modulation through the Fourier transform:

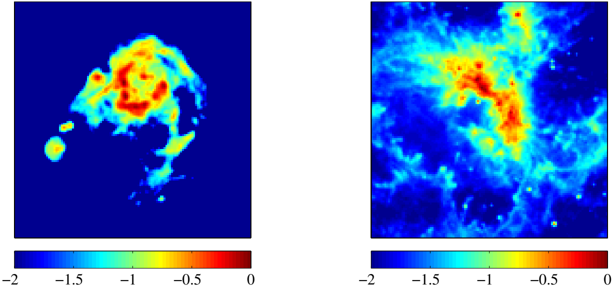
$$\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\,\mathbf{F}\,\mathbf{A} .$$

- Naively, expressing the application of the w-modulation in this manner is computationally less efficient that the original formulation but it has two important advantages.
- Different w for each (u, v), while still exploiting FFT.
- Many of the elements of $\hat{\mathbf{C}}$ will be close to zero.
- Support of w-modulation in Fourier space determined dynamically.

Spread spectrum effect for varying w

Results on simulations

- Perform simulations to assess the effectiveness of the spread spectrum effect in the presence of varying w.
- Consider idealised simulations with uniformly random visibility sampling.

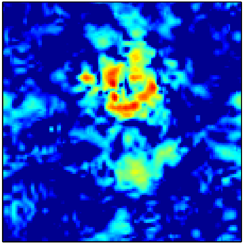


(a) M31 (b) 30Dor

Figure: Ground truth images in logarithmic scale.

Spread spectrum effect for varying w

Results on simulations



(a) $w_{\rm d} = 0 \rightarrow$ SNR$= 5$dB

Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

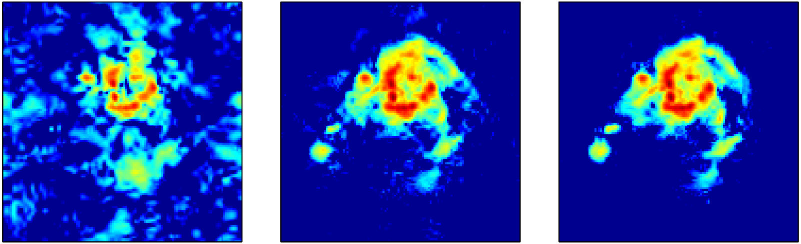
(a) $w_{\mathrm{d}} = 0 \rightarrow$ SNR$=$ 5dB

(c) $w_{\mathrm{d}} = 1 \rightarrow$ SNR$=$ 19dB

Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

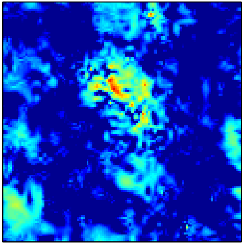


(a) $w_\mathrm{d} = 0 \rightarrow$ SNR$=$ 5dB (b) $w_\mathrm{d} \sim \mathcal{U}(0, 1) \rightarrow$ SNR$=$ 16dB (c) $w_\mathrm{d} = 1 \rightarrow$ SNR$=$ 19dB

Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

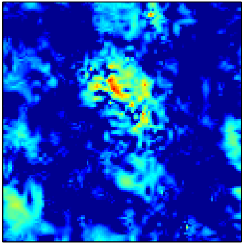


(a) $w_\mathrm{d} = 0 \rightarrow$ SNR$=$ 2dB

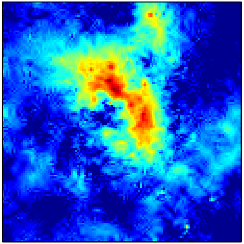
Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations



(a) $w_{\rm d} = 0 \rightarrow$ SNR$=$ 2dB

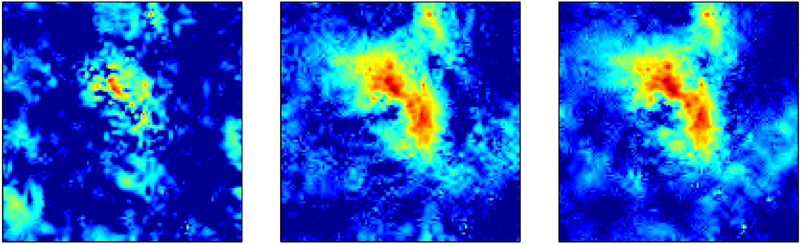


(c) $w_{\rm d} = 1 \rightarrow$ SNR$=$ 15dB

Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

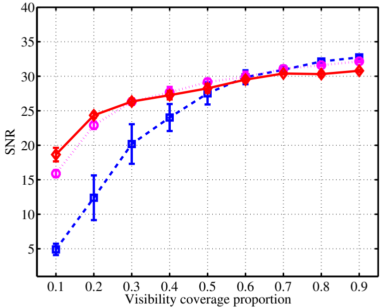


(a) $w_{\mathrm{d}} = 0 \rightarrow$ SNR$=$ 2dB (b) $w_{\mathrm{d}} \sim \mathcal{U}(0, 1) \rightarrow$ SNR$=$ 12dB (c) $w_{\mathrm{d}} = 1 \rightarrow$ SNR$=$ 15dB

Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets

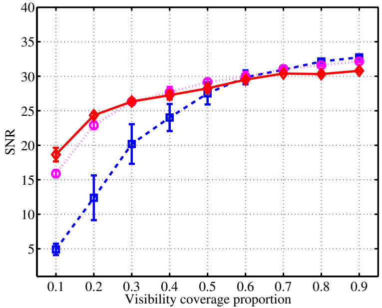
Figure: Reconstruction fidelity for M31.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w.

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets

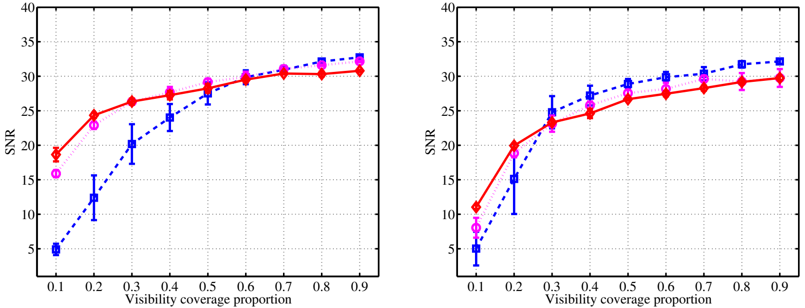
Figure: Reconstruction fidelity for M31.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w.

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets (b) Dirac basis

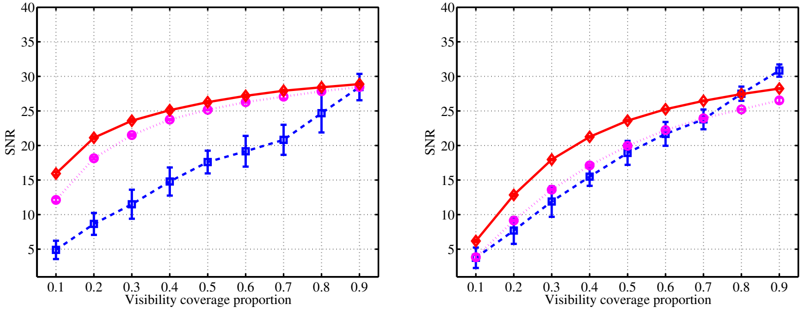
Figure: Reconstruction fidelity for M31.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w.

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets (b) Dirac basis

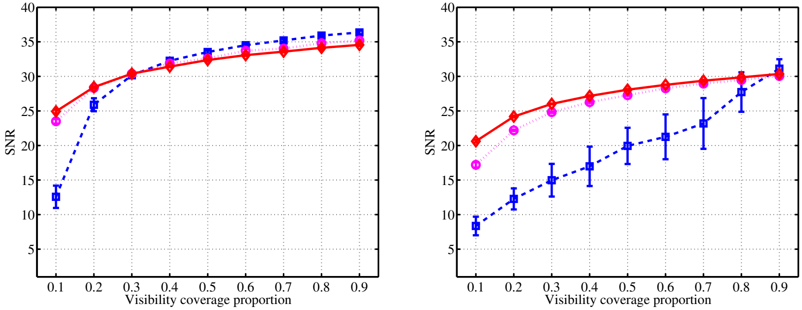
Figure: Reconstruction fidelity for 30Dor.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w.

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) M31 (b) 30 Dor

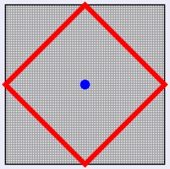
Figure: Reconstruction fidelity using SARA.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w.

- As expected, for the case where coherence is already optimal, there is little improvement.

Public codes

SOPT code http://basp-group.github.io/sopt/



Sparse OPTimisation
Carrillo, McEwen, Wiaux

SOPT is an open-source code that provides functionality to perform sparse optimisation using state-of-the-art convex optimisation algorithms.

PURIFY code http://basp-group.github.io/purify/

Next-generation radio interferometric imaging
Carrillo, McEwen, Wiaux

PURIFY is an open-source code that provides functionality to perform radio interferometric imaging, leveraging recent developments in the field of compressive sensing and convex optimisation.

Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.

Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.

Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.

Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.